

Chapter 2. DATA FILES AND STRUCTURES

The MRFSS data bases include:

Raw data files;

- ! creel intercept interviews, and
- ! telephone interviews,

Summary files;

- ! non-fishing/fishing households (from the telephone survey)
- ! trip estimates, and
- ! catch estimates.

The raw data files have multiple record types in the data base. The summary files have only one record type.

Creel Intercept Interview Files

There are six record types in the creel intercept interview data bases. For every interview, there is a unique ID_CODE coded on all records associated with that interview. The ID_CODE is used to link all records of an interview. The ID_CODE consists of survey type, interviewer number, date of the survey, and interview number. Interview numbers begin with one and are assigned consecutively to interviews within a site/date combination.

Record Types

- ! Type 1 - Angler/Trip Data
 - ! I1 records contain fisherman and trip information including ID_CODE, location of the interview, fishing area and mode, hours fished, gear used, geographic residence and avidity of the angler, species targeted, presence/absence of catch, number of contributors to the catch, and total numbers of I2-I6 records associated with the interview.
 - ! Special questions such as artificial reef fishing, sea turtle sightings, tournament fishing, etc. vary by year and area, and are usually recorded on the I1 records.
 - ! There is only one I1 record per interview.

! Type 2 - Unavailable Catch (Catch Type B1 and B2)

! I2 records contain the ID_CODE, and the number and disposition of unavailable catch by species. Disposition codes categorize catch as released alive (catch type B2), and used for bait, filleted, or other consumptive use (catch type B1). For unavailable catch, the species name is reported by the angler but can not be verified by the interviewer.

! There can be multiple I2 records for each interview, one record for each species code/disposition combination. More than one disposition code can be used for a particular species code.

! The variable containing the numbers of fish caught (NUM_FISH) is additive across all records.

! In certain years and areas, shellfish catch information was collected and coded on I2 records. These records are kept in separate data files.

! Type 3 - Available Catch (Catch Type A)

! I3 records contain the ID_CODE, species code, numbers of fish, disposition of the catch, and length and weight measurements for fish brought to land and identified to species by the interviewer (catch type A).

! There can be multiple I3 records for an interview, one record for each individually weighed/measured fish. If no fish were weighed or measured, there will be one record for a species. Only one disposition code can be used for each species.

! The variable containing the number of fish caught (FISHINSP) is not additive across records. This number is the total number of that species caught and is carried across all records for that species with individual length/weight measurements.

! In certain years and areas, shellfish catch information was collected and coded on I3 records. These records are kept in separate data files.

! Type 4 - Group Catches

! I4 records contain the ID_CODE of the interviewed angler, as well as the ID_CODE of the angler whose records contain combined type A catches.

! There is only one I4 record per interview and this occurs only when A-type catches can not be separated for individual anglers. If catches can be separated, no I4 record exists.

! Type 5 - Socio-Economic Data

! GP5 records contain the ID_CODE and socio-economic data collected in 1983 on the Pacific coast, in 1987 nation-wide, and in 1990 on the Atlantic and Gulf coasts.

! Questions and resulting data vary between the three years. Travel costs were the primary emphasis.

! There is only one GP5 record per interview.

! Type 6 - First Interview of Fishing Party

! I6 records contain the ID_CODE of the interviewed angler, and the ID_CODE of the first individual interviewed within a fishing party.

! This record was begun in 1991 in order to link all interviews within a party together.

! There is only one I6 record per interview.

File Naming Conventions

The intercept SAS file naming conventions are standard across years and regions (Table 2).

Table 2. Naming conventions for SAS intercept files from the MRFSS. YY = year; W = wave; ST = state.

Record Type	Coastwide Sets	State Sets
1	I1_YYW.SSD	ST1_YYW.SSD
2	I2_YYW.SSD	ST2_YYW.SSD
3	I3_YYW.SSD	ST3_YYW.SSD
4	I4_YYW.SSD	ST4_YYW.SSD
5	I5_YYW.SSD	ST5_YYW.SSD
6	I6_YYW.SSD	ST6_YYW.SSD

The naming conventions for the ASCII intercept files are: Atlantic/Gulf coast datasets are named AG_YYW.INT, and state sets are named STYYW.INT. YY is year, W is wave, and ST is state. In the ASCII files, all record types are kept together in one file per wave. For each angler

interview, the type 1 record will be followed by all the type 2 records, then the type 3 records, etc. The first column contains the record type.

Fishing Household Telephone Interview Files

There are three record types in the fishing household telephone interview files. The record types for the telephone fishing household interview files are similar to the intercept interviews, in that a distinct CODENUM is included in every record associated with a specific household. CODENUM is a unique number for each household dialed during a wave.

Record Types

! T1 - Household Information

! T1 records contain fishing household information including CODENUM, state and county of residence, number of anglers who went fishing in the last 12 months and the last 2 months, number of interviewed anglers, number who refused to be interviewed, number who were unavailable, and the number otherwise not interviewed due to language barriers, age, death, etc.

! Special questions such as number of shellfishing or spiny lobster fishermen (12-month and 2-month) exist on the T1 records for certain years.

! There is only one I1 record per fishing household.

! T2 - Angler Information

! T2 records contain information for each interviewed angler including an id number for each angler within a household, whether the information was provided by the angler or by someone else, language, whether the angler could recall all trips in the 2-month period, and the total number of trips taken by the angler.

! T2 records for 1993 include tournament participation data for Connecticut.

! There is one T2 record for each interviewed angler in a fishing household.

! T3 - Fishing Trip Information

! T3 records contain trip information including CODENUM, the angler id code, a consecutive number for each trip taken by the angler, date of the trip, fishing mode, gear used, area fished, boat access site characteristics, and state and county where the trip occurred.

- ! There is one T3 record for each trip taken by an interviewed angler, unless the angler cannot remember key details of the trip. In that case, the total number of T3 records may be less than the total number of trips reported by an angler on the T2 record.

- ! T3 records for some years and areas contain special trip information on spiny lobster trips, numbers of striped bass caught and kept, tournament trip data, oil platform/artificial reef use, etc.

File Naming Conventions

The telephone SAS file naming conventions are standard across regions and years (Table 3).

Table 3. Naming conventions for SAS telephone files from the MRFSS. YY = year; W = wave; ST = state.

Record Type	Atlantic and Gulf Coast Sets	State Sets
1	T1_YYW.SSD	ST1_YYW.SSD
2	T2_YYW.SSD	ST2_YYW.SSD
3	T3_YYW.SSD	ST3_YYW.SSD

The naming convention for ASCII telephone files is: Atlantic/Gulf Coast files are named WAVEX_YY.HSE, and state files are named STYYW.HSE. All record types are kept together in one file per wave. For each household, the type 1 record will be followed by all the type 2 records, then the type 3 records, etc. The first column contains the record type.

Non-Fishing Household Files

Non-fishing household files are derived from census data and telephone interview data. They contain the number of sampled fishing and non-fishing households, and the total number of households in the county for each coastal county by wave. There is one record per county. The naming convention is NFYYW.SSD for SAS files and NFYYW.DTA for ASCII files. These files are not currently separated into state files and are not generally distributed but can be made available upon request.

The primary use of these files is in the estimation procedure; however, they should be used for proper weighting when calculating prevalence rates (% of households that fish). The sampling procedure weights counties by the square root of the county population to ensure that rural, less populated areas have at least some representation. Therefore, when the raw data are used to calculate prevalence rates, they should be reweighted, or rural county activity levels will be given too much weight.

Catch Estimate Files

The catch files have one record for each strata (state, wave, fishing mode, and fishing area) and species combination. Each catch file record contains the estimates for each of the three catch types (A, B1, and B2), their variances, mean weights and variances, sample size information, and other variables and variances used to calculate the estimates.

Note: Catch records also contain trip estimates by strata/species. Care must be taken using these estimates for trips since there are multiple species records in each cell with duplicate trip records.

The naming conventions for catch files are AG_YYW.SSD and ST_YYW.SSD for Atlantic/Gulf coast and state SAS datasets, respectively; and AG_YYW.EST and ST_YYW.EST for ASCII datasets.

Trip Estimate Files

Trip files have one record for each state, wave, and fishing mode combination. Each trip record then has trip estimates for the separate fishing areas. Trip files also contain adjustment ratios such as non-coastal residents, households without telephones, sample sizes, and variance estimates for trips and other variables used for estimation procedures.

Naming conventions for trip files are AG_YYW.SSD and ST_YYW.SSD for Atlantic/Gulf Coast and state SAS datasets, respectively; and AG_YYW.TRP and ST_YYW.TRP for ASCII datasets for the Atlantic/Gulf coasts and the separate states, respectively.

Appendix B contains variable names and codes, and presence/absence of variables from 1979-1994 for the intercept and telephone interview data. Appendix C lists estimate file variables.